Notes on Measure-Theoretic Probability

Alex Yakubovich
Department of Mathematics & Statistics
York University

September 29, 2011

1 Basics of Measure Theory

We want to define functions that measure the size of subsets of a given set. We are familiar with a few such functions - length of intervals of $\mathbb{R}$, area of of regions in $\mathbb{R}^2$, and volume of solids in $\mathbb{R}^3$. We would like to generalize to more arbitrary sets.

It turns out we cannot construct a measure on the power set on $\mathbb{R}$ which extends basic properties of the measures listed above, such as translation invariance.

Therefore, we will only look at certain subsets - members of sigma algebras.

A σ **-Algebra** $\mathcal{F}$ on a set Ω is a collection of subsets of Ω satisfying:

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The sigma algebra generated by a set A is the smallest sigma algebra containing A. It is defined as the intersection of all sigma algebras containing A, which can easily be verified to be a sigma algebra.

The sigma algebra most relevant for probability theory is the Borel sigma algebra, which is generated by open sets:

$$\mathcal{B}^k := \sigma(\mathcal{O}(\mathbb{R}^k))$$

Sets in a sigma algebra are called **measurable sets**. Next, we define the notion of a measure on such sets.

Given a **measurable space** $(\Omega, \mathcal{F})$, a function $f : \mathcal{F} \to \mathbb{R}$ is called a **measure** on Ω iff:

1. $0 \leq \mu(A) \leq \infty$
2. $\mu(\emptyset) = 0$

3. $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \quad \forall A_n \in \mathcal{F}$ disjoint

$(\Omega, \mathcal{F}, \mu)$ is a **measure space**.

If $\mu(\Omega) = 1$, then μ is a **probability measure**, usually denoted $p(.)$, and $(\Omega, \mathcal{F}, p)$ is a **probability space**.

Measures extends the following properties of length:

1. (Monotonicity) $A \subset B \implies \mu(A) \leq \mu(B)$
2. (Subadditivity) $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$
3. (Continuity) If $A_1 \subset A_2 \ldots$ (or $A_1 \supset A_2 \ldots$), then $\mu\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$ where $\lim_{n\to\infty} \mu(A_n) = \bigcup_{n=1}^{\infty} A_n$ (or $\bigcap_{n=1}^{\infty} A_n$)

2 Cumulative Distribution Function

Given a probability space $(\mathbb{R}, \mathcal{B}, P)$, the **Cumulative Distribution Function** of P is the function $F : \mathbb{R} \to [0, 1]$ defined by:

$$F(x) = P((-\infty, x]) \quad \forall x \in \mathbb{R}$$

Properties:

1. $\lim_{x\to-\infty} F(x) = 0$
2. $\lim_{x\to\infty} F(x) = 1$
3. F nondecreasing
4. F right-continuous

3 Product Measure

Sometimes we want to measure the size of subsets of a cross product of two sets, for example, a region $R \subset \mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

To take the measure, we need a sigma-algebra. Unfortunately, the cross product of sigma algebras is not generally a sigma algebra, so instead, we take the sigma algebra generated by the product of sigma algebras. Thus:

Given sets $(\Omega_i, \mathcal{F}_i) \quad i = 1 \ldots n$, we build the measurable space:

$$\left(\prod_{i=1}^{n} \Omega_i, \sigma\left(\prod_{i=1}^{n} \mathcal{F}_i\right)\right)$$

The next question is whether measures distribute across products of sets like it does across disjoint unions. We will need to restrict our discussion to sigma-finite measures - measures that are finite on every member of a countable decomposition of the space:

The next question is whether measures distribute across products of sets like it does across disjoint unions. We will need to restrict our discussion to sigma-finite measures - measures that are finite on every member of a countable decomposition of the space:

A measure μ on $(\Omega, \mathcal{F})$, is σ **-finite** iff $\exists \{A_i\}_{i=1}^{\infty}$ s.t:

1. $\Omega = \bigcup_{i=1}^{\infty} A_i$
2. $\mu(A_i) < \infty \quad \forall i \in \mathbb{N}$

For example, the Lebesgue Measure is sigma-finite on R, since it is finite on every interval $(-n, n)$ and $\mathbb{R}$ is exhausted by countably many such intervals.

Another example is the counting measure, which is sigma-finite iff the space is sigma-finite.

Given $\{(\Omega_i, \mathcal{F}_i, \mu_i)\}_{i=1}^{n}$, where the μ_i are σ-finite measures, there exists a unique measure $\mu_1 \times \ldots \times \mu_n$ on $(\prod_{i=1}^{n} \Omega_i, \sigma(\prod_{i=1}^{n} \mathcal{F}_i))$, called the the **product measure**, which satisfies:

$$\prod_i \mu_i \left(\prod_i A_i \right) = \prod_i \mu_i(A_i)$$

4 Joint Cumulative Distribution Function

Now that are we able to measure subsets of a product space, we can define a CDF associated with the product measure:

Given a probability space $(\mathbb{R}^k, \mathcal{B}^k, p)$, the **joint CDF** of p is the function $F : \mathbb{R}^k \to [0, 1]$ defined by:

$$F(x_1, \ldots, x_k) = p((-\infty, x_1] \times \ldots \times (-\infty, x_2]), \quad x_i \in \mathbb{R}$$

It turns out that there is a 1:1 correspondence between probability measures on the above space and joint CDFs on $\mathbb{R}^k$ - a given probability measure defines a joint CDF, and a joint CDF corresponds to a unique probability measure.

Given a joint CDF, we can find the marginal CDFs:

$$F_1(x) = \lim_{x_j \to \infty \quad \forall j \neq i} F(x_1, \ldots, x_{i-1}, x_{i+1} \ldots x_k)$$

Given the marginal CDFs, we generally can't find the joint CDF.

A special case is when the joint CDF factors into a product of the marginals:

$$F(x_1, \ldots x_n) = F(x_1)F(x_2) \ldots F(x_n) \quad \forall (x_1, \ldots x_n) \in \mathbb{R}^k$$

In this case , the probability measure associated with F is the product measure.

5 Inverses of Set Functions

Given $f : \Omega \to \Lambda$, and a set $B \subset \Lambda$, the inverse image of B under f is:

$$f^{-1}(B) = \{w \in \Omega : f(w) \in B\} \equiv \{f \in B\}$$

If $|f^{-1}(y)| = 1 \quad \forall y \in B$, then f^{-1} is a function and f is injective.

Inverse functions commute with complements/unions:

1. $f^{-1}(B^c) = (f^{-1}(B))^c \quad \forall B \in \Lambda$
2. $f^{-1}(\cup B_i) = \cup f^{-1}(B_i) \quad \forall B \in \Lambda$

6 Measurable Functions

Let $(\Omega, \mathcal{F}), (\Lambda, \mathcal{G})$ be measurable spaces and $f : \Omega \to \Lambda$
f is a **measurable function** from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ iff $f^{-1}(\mathcal{G}) \subset \mathcal{F}$

A **random variable (aka Borel function)** is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$

$f^{-1}(\mathcal{G})$ is a sub-sigma field of $\mathcal{G}$. It is the σ **field generated by** f

Now that we have defined what a random variable is, we show that there enough functions that meet the definition of RV. In particular, instead of requiring that pullbacks of ANY borel set are measurable sets, it suffices to look at open right-infinite intervals. Furthermore, given two borel functions, we can construct others through linear combinations, products, quotients, limits and compositions. We also establish the connection between continuous functions and Borel functions - continuous mappings of Borel sets into R are measurable.

Let $(\Omega, \mathcal{F})$ be a measurable space. Then:

1. f Borel $\iff f^{-1}(a, \infty) \in \mathcal{F}$
2. f, g Borel $\Rightarrow fg, af + bg, f/g$ Borel, where $a, b \in \mathbb{R}$ and $g(\omega) \neq 0 \quad \forall \omega \in \Omega$
3. $\{f_n\}_{n=1}^{\infty}$ Borel $\Rightarrow \sup f_n, \inf f_n, \limsup f_n, \limsup f_n$ Borel, provided these limits exist. Otherwise, if we define $A = \{\omega \in \Omega : \lim_{n\to\infty} f_n(\omega)$ exists $\}$ $\in \mathcal{F}$, then the following function is Borel:

$$h(\omega) = \begin{cases} \lim_{n\to\infty} f_n(\omega), & \text{if } \omega \in A \\ f_1(\omega) & \text{else} \end{cases}$$

4. If f is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ and g is measurable from $(\Lambda, \mathcal{G})$ to $(\Delta, \mathcal{H})$, then $g \circ f$ is measurable from $(\Omega, \mathcal{F})$ to $(\Delta, \mathcal{H})$.
5. Let Ω be a Borel set in $\mathbb{R}^p$. Then if $f : \Omega \to \mathbb{R}^q$ is continuous, then f is measurable.

7 Cumulative Distribution Functions

We have defined CDFs as real valued functions associated with measures. We now define CDFS associated with measurable functions (RV), as well as distributions of RV:

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and f be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.

f **induces a measure** $\mu \circ f^{-1}$ on $\mathcal{G}$, defined as:

$$\mu \circ f^{-1}(B) = \mu(f^{-1}(B)), \quad \forall B \in \mathcal{G}$$

If $\mu = p$ a probability measure, and $f = X$ a random variable, then $P \circ X^{-1}$ is the distribution/law of X, denoted P_x. The cdf associated with the measure P_x is the c.d.f of X, denoted F_X

8 Simple Functions

A **simple function** $\phi : \Omega \to \mathbb{R}$ has the form: $\phi(\omega) = \sum_{i=1}^{n} a_i \mathbb{I}_{A_i}(\omega)$ where $a_i \in \mathbb{R}$ and A_i are measurable sets.

Properties of simple functions:

1. $\sigma(\phi) = \sigma(\{A_1, \ldots, A_n\})$
2. $\phi \geq 0 \iff a_i \geq 0 \quad i = 1 \ldots n$

9 Integration

We define the integral of a simple function as a linear combination of the measures of the sets its defined over. Integrals of Borel functions are defined as supremums over integrals of lesser simple functions.

To deal with functions taking on negative values, we decompose a function into the difference of its non-negative and negative values, and integrate each component using the above definition.

To integrate a measurable function over a set, we integrate the product of the function with the set's indicator function over the whole space.

1. $\displaystyle\int_{\Omega} \phi d\mu := \sum_{i=1}^{n} a_i \mu(A_i)$

2. Let f be a non-negative Borel function and

$$\mathcal{S}_f := \{\phi \text{ a simple fn s.t } \phi(\omega) \leq f(\omega) \quad \forall \omega \in \Omega\}$$

Then the integral of f w.r.t μ is:

$$\int_{\Omega} f d\mu := \sup_{\phi \in \mathcal{S}_f} \int_{\Omega} \phi d\mu$$

$\Rightarrow$ For any Borel function f, there exists a sequence $\phi_1 \dots \phi_n$ of simple functions s.t. $0 \leq \phi_n \leq f \quad \forall n$ and:

$$\lim_{n \to \infty} \phi_n d\mu = \int f d\mu$$

3. Any Borel function f can be written as: $f(\omega) = f_+(\omega) + f_-(\omega)$, where

$$f_+(\omega) = \max\{f(\omega), 0\} \geq 0$$

$$f_-(\omega) = \max\{-f(\omega), 0\} \leq 0$$

   $\int_f d\mu$ exists iff one of $\int f_+ d\mu, \int f_- d\mu$ is finite. It is defined as:

$$\int_f d\mu = \int f_+ d\mu - \int f_- d\mu$$

4. If both $\int f_+ d\mu, \int f_- d\mu < \infty$, then f is integrable. The integral of f over a measurable set A w.r.t measure μ is defined as:

$$\int_A f d\mu := \int_\Omega \mathbb{I}_A f$$

10 Notation for Integrals in Probaiblity Theory

Given a R.V. X from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B})$, the **expectation** of X is:

$$\mathbb{E}(X) := \int_\Omega X dP$$

If F is the C.D.F of probability measure P on $(\mathbb{R}^k, \mathcal{B}^k)$, then

$$\int_\Omega f(x) dF := \int_\Omega f(x) dP$$

11 Properties of Integrals of Measurable Functions

1. Integration is a linear operator
2. Given $(\Omega, \mathcal{F}, \mu)$ and f, g Borel functions:
   (a) $f \leq g \quad$ a.e $\Rightarrow \int f d\mu \leq \int g d\mu \quad$ (*Monotonicity*)
   (b) $f \geq 0, \int f d\mu = 0 \Longrightarrow f = 0$ a.e
3. $\left|\int f d\mu\right| \leq \int |f| d\mu \quad$ (*Triangle Inequality*)
4. $f = g \quad$ a.e $\Longrightarrow \int f d\mu = \int g d\mu$
5. $f \geq 0 \quad$ a.e $\Longrightarrow \int f d\mu \geq 0$

12 The Lebesgue Measure / Integral

The measure which motivated Measure Theory is the Lebesgue Measure - the unique measure on subsets of sigma algebras of Rn which agrees with length for intervals, and shares its properties when it comes to more arbitrary subsets.

As the Lebesgue measure generalizes the notion of length, the Lebesgue integral generalizes the Reiman integral. The two agree on Rieman integrable functions (continuous, bounded functions over compact sets), but the Lebesgue integral is defined for all Lebsgue-measurable functions, which may not be Reimann integrable.

Given a a random variable X on probability space $(\Omega, \mathcal{F}, P)$, we can construct a probability measure on $(\mathbb{R}, \mathcal{F})$ called the distribution of X and a CDF associated with X (=CDF of the measure)

Random variables provide a way to use measures on arbitrary sets to measure sets in $\mathbb{R}$(Borel sets). There are two functions associated with a random variable - a distribution, which measures any borel sets, and a CDF, which measures sets $(-\infty, x)$.

13 Convergence Theorems

Next, we answer the question of when a limit and an integral can be interchanged. There are 3 results providing sufficient conditions for this:

Let $f_1, f_2 \ldots$ be Borel functions(μ-measurable functions from $(\Omega, \mathcal{F})$ to $(\mathbb{R}), \mathcal{B})$. Then:

1. *(Fatou's Lemma)* $f_n \geq 0 \Longrightarrow \liminf \int f_n \leq \int \liminf f_n$
2. *(Dominated Convergence Theorem)* IF $\lim_{n\to\infty} f_n = f \quad$ a.e and $\exists g$ integrable s.t. $|f_n| \leq g \quad$ a.e , THEN $\lim \int f d\mu = \int \lim f d\mu$
3. *(Monotone Convergence Theorem)* IF $0 \leq f_1 \leq f_2 \leq \ldots$ and $\lim_{n\to\infty} f_n = f \quad$ a.e , THEN $\lim \int f d\mu = \int \lim f d\mu$

14 Change of Variables Theorem

The Change of Variables Theorem defines the integral of a composition of a measurable function and a borel function:

Let f be measurable from $(\Omega, \mathcal{F}, \mu)$ to $(\Lambda, \mathcal{G})$ and g be measurable from $(\Lambda, \mathcal{G})$ to $(\mathbb{R}, \mathcal{B})$. Then:

$$\int_\Omega g \circ f d\mu := \int_\Omega g d(\mu \circ f^{-1})$$

15 Fubini's Theorem

Let μ_i be sigma-finite measures on $(\Omega_i, \mathcal{F}_i)$ and f a borel function on $\prod(\Omega_i, \mathcal{F}_i)$, and either $f \geq 0$ or $\int |f| d\mu_1 \times \mu_2 < \infty$.
Then $g(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1$ exists and is a Borel function on Ω_2 whose integral w.r.t μ_2 exists and

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d\mu_1 \times \mu_2 = \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1 \right] d\mu_2$$

16 The Radon-Nikodym Derivative

Let μ, λ be measures on $(\Omega, \mathcal{F})$. λ is *absolutely continuous* w.r.t μ, denoted $\lambda << \mu$, if

$$\mu(A) = 0 \Longrightarrow \lambda(A) = 0 \quad \forall A \in \mathcal{F}$$

Given μ, we can construct an absolutely continuous measure λ by setting:

$$\mu(A) := \int_A f d\mu \quad \forall A \in \mathcal{F}$$

where f is a non-negative Borel function. It turns out that under some assumptions, every $\lambda << \mu$ can be uniquely written this way.

(*Radon-Nikodym Theorem*) Suppose μ is a σ-finite measure on $(\Omega, \mathcal{F})$, and $\lambda << \mu$ is also a measure on $(\Omega, \mathcal{F})$. Then there exists a non-negative Borel function f on Ω s.t:

$$\lambda(A) = \int_A f d\mu \quad \forall A \in \mathcal{F}$$

f is called the Radon-Nikodym derivative of μ, denoted $\dfrac{d\lambda}{d\mu}$ and is unique a.e μ

When $\lambda(\Omega) = \int_\Omega f d\mu = 1, \lambda$ is a probability measure, and f is its *probability density function (p.d.f).*

When λ is the probability measure induced by a R.V X ($\lambda = P \circ X^{-1}$), f is the p.d.f of X (or of F_X)

(Discrete CDF) Let $a_1 < a_2 < \ldots$ and and $p_1, p_2 \ldots$ be sequences in $\mathbb{R}$, with $\sum_i p_i = 1$. We can define a c.d.f $F : \mathbb{R} \to [0, 1]$ as follows:

$$F(x) = \begin{cases} \sum_{i=1}^n p_i & \text{if } x \in [a_n, a_n + 1) \\ 0 & \text{otherwise} \end{cases}$$

F is associated with a probability measure P on $(\mathbb{R}, \mathcal{B})$.
Let μ be the counting measure.

Then $P << \mu$, and by R-N theorem:

$$P(A) = \sum_{i\in\mathcal{I}} p_i = \sum_{i\in\mathcal{I}} f(a_i), \text{where } \mathcal{I} = \{i : A \cap [a_i, a_{i+1}) \neq \emptyset\}$$

$$\frac{dP}{d\mu} = f(\omega) = \begin{cases} p_i & \text{if } w = a_i \\ 0 & \text{otherwise} \end{cases}$$

Let F a CDF differentiable in the calculus sense and $f = F'$. From FTC:

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

Let P be the probability measure on $(\mathbb{R}, \mathcal{B})$ associated with F. Then by Radon-Nikodym Theorem on lebesgue measure μ, with $P << \mu$, there exists non-negative Borel function f s.t.:

$$P(A) = \int_A f d\mu \quad \forall A \in \mathcal{B}$$

$f = \dfrac{dP}{d\mu}$ is the *p.d.f* of F w.r.t. Lebesgue measure μ. It agrees with the *p.d.f* obtained by differentiating F using calculus.

We can relax the assumption that F is differentiable - it need only be *absolutely continuous*.

A function F is absolutely continuous on $\mathbb{R}$ if $\forall \epsilon > 0, \exists \delta > 0$ such that $\forall \{(a_i, b_i)\}_{i=1}^n$ finite collections of disjoint, bounded open intervals,

$$\sum |b_i - a_i| < \delta \Longrightarrow \sum |F(b_i) - F(a_i)| < \epsilon$$

A function is $\mu-$differentiable $\iff$ it is absolutely continuous on $\mathbb{R}$

17 Moments and their Properties

Properties of 1st/2nd Moment:

1. $Cov(X_i, X_j) \leq Var(X_i)Var(X_j)$
2. Variance matrix is non-negative definite ($y^t Var(X) y \geq 0 \quad \forall y$)
3. Independence $\Longrightarrow$ Correlation
4. $Y = c^T X \Longrightarrow \mathbb{E}[Y] = c^T \mathbb{E}[X]$ and $Var(E) = c^T Var(X) c$

Three useful inequalities:

1. (*Cauchy-Schwartz Inequality*): $[\mathbb{E}(XY)]^2 \leq [\mathbb{E}(X)\mathbb{E}(Y)]^2$, for R.V X, Y
2. *(Jensen's Inequality:)* $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$, convex function f and $R.V.X$

3. *(Chebyshev's Inequality)* Let X be a RV and $\phi : \mathbb{R} \to [0, \infty)$ nondecreasing, even function. Then

$$\phi(t)P(|X| \geq t) \leq \int_{\{|X|\geq t\}} \phi(X)dP \leq \mathbb{E}[\phi(X)], \quad \forall t \geq 0$$

Given a random vector $X \in \mathbb{R}^k$:

1. The *Moment Generating Function* of X (or P_X) is $\Psi : \mathbb{R} \to \mathbb{R}$ given by:

$$\Psi_x(t) = \mathbb{E}[e^{t^T X}], \quad t \in \mathbb{R}^k$$

2. The *Characteristic Function* of X (or P_X) is $\phi : \mathbb{R} \to \mathbb{C}$ given by:

$$\phi_X(t) = \mathbb{E}[e^{it^T X}], \quad t \in \mathbb{R}^k$$

3. The *Cumulant Generating Function* of X (or P_X) is:

$$\kappa_X(t) = \log \Psi_X(t), \quad \text{if } 0 < \Psi_X(t) < \infty$$

18 Properties of m.g.f and c.h.f

1. If mgf is finite in a neighbourhood of 0, then all moments exist

2. $y = A^T X + c$
   $\implies \Psi(u) = \mathbb{E}[e^{u^T(A^T X + c)}] = e^{c^T u}\Psi(Au)$
   $\implies \phi(u) = \mathbb{E}[e^{iu^T(A^T X + c)}] = e^{ic^T u}\phi(Au)$

3. $X_1 \dots X_k$ are independent and $Y = \sum_{i=1}^k X_i$, then:

$$\Psi_Y(t) = \prod_{i=1}^k \psi_{X_i}(t) \text{ and } \phi_Y(t) = \prod_{i=1}^k \phi_{X_i}(t)$$

4. *(Uniqueness of mgf/chf).* Suppose X, Y are random vectors. Then:
   (a) $\phi_X(t) = \phi_Y(t) \quad \forall t \in \mathbb{R}^k \implies P_X = P_Y$
   (b) $\psi_X(t) = \psi_Y(t) < \infty \quad \forall t$ in a neighborhood of $0 \implies P_X = P_Y$

5. *(Symmetry)* (A random vector X is symmetric about 0 iff X and $-X$ have the same distribution.)

$$X \text{ is symmetric about } 0 \iff \phi_X \text{is real valued}$$

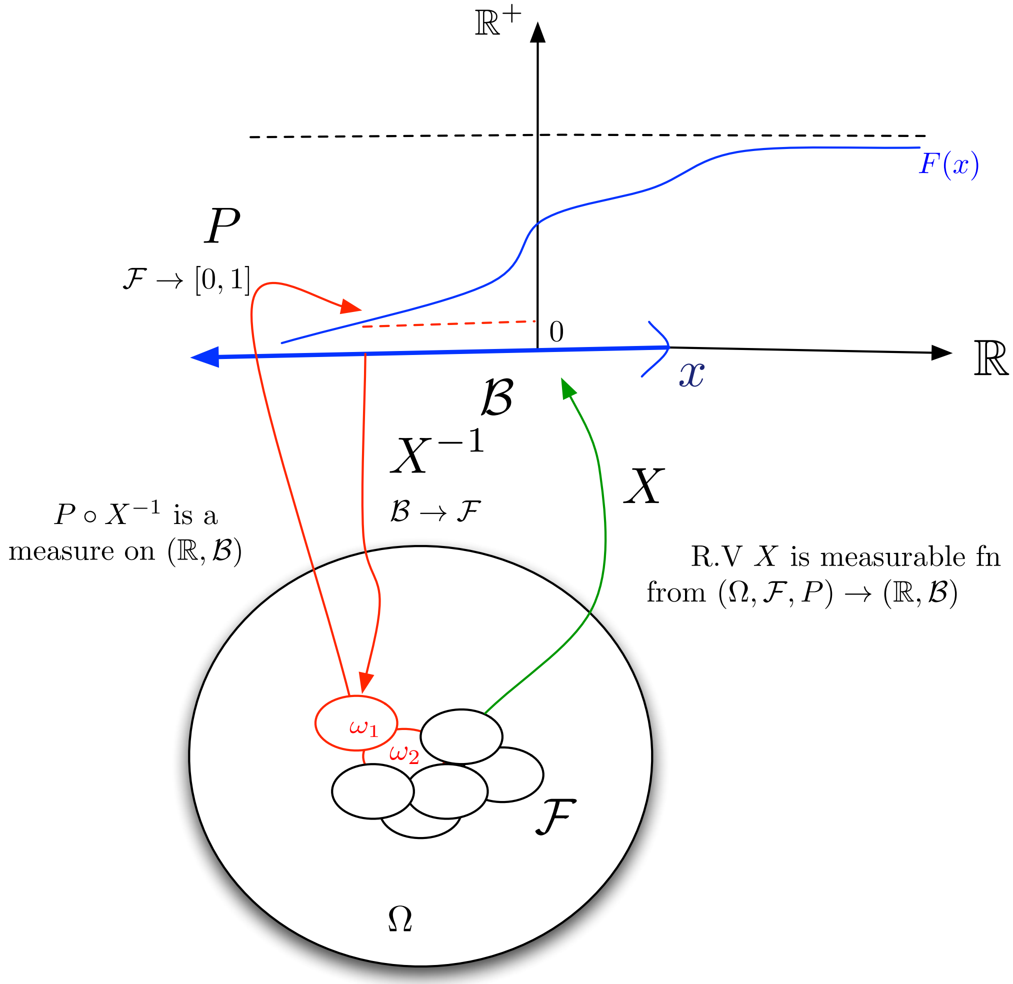

Figure 1: Mechanics of a random variable